Random and Quasirandom Graphs	Instructor: Padraic Bartlett
Lecture 1: Background: Probability, $o(n)$, and Connecting LA to GT	
Week 4	*Mathcamp 2012*

This lecture is a brief crash course in probabilistic notation and ideas, $o(n)$ notation, and the connection between graph theory and linear algebra; consequently, it's going to be definition-heavy and full of ideas/vocabulary rather than theorems. But, if we do this, we can actually survive the next n lectures!

1 Probability

Definition. A **probability space** is a set X together with a "likelihood" or "probability" function $\mathbb{P}$, that takes in subsets of X and outputs numbers in $\mathbb{R}$. We ask that $\mathbb{P}$ satisfies the following three properties:

- For any $A \subseteq X$, $0 \leq \mathbb{P} \leq 1$.
- $\mathbb{P}(X) = 1$.
- For any $A, B \subseteq X$, if A and B are disjoint, then $\mathbb{P}(A) + \mathbb{P}(B) = \mathbb{P}(A \cup B)$.

We will often call subsets of X **events**: when we look at $\mathbb{P}(X)$, we are thinking of this as the "likelihood that X" happens" according to our "likelihood" function P.

Given a function $f : X \to \mathbb{R}$, if X is finite, we denote the **expected value of** f as the sum

$$\mathbb{E}(f) = \sum_{x \in X} f(x) \cdot \mathbb{P}(x).$$

(For infinite sets X, we would use the integral of $f(x)$ over X, where the measure we're using on X is the function $\mathbb{P}$. If you don't know what a measure is, or even if you do, completely ignore this note.)

From these properties, we can deduce a number of other pretty useful-straightforward properties:

Proposition. Take any probability space $(X, \mathbb{P})$.

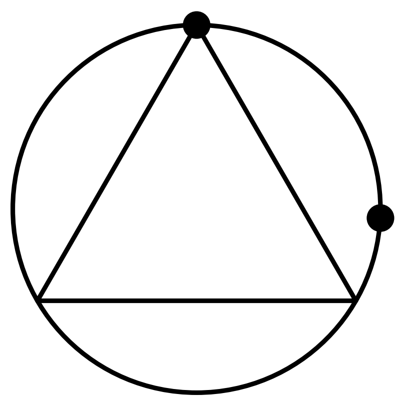
1. $\mathbb{P}(A^C) = 1 - \mathbb{P}(A)$.
2. $\mathbb{P}(\emptyset) = 0$
3. Given any two events A, B, $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.

These properties are fairly straightforward, and you should prove them if you don't believe them.

In mathematics, people are rarely precise enough to always explicitly state the probability space they're working in; rather, they'll just say "if you take a random *blah*, it will have property *blarg* with probability *bargle*."[1] This is sometimes not a good idea, as the following example illustrates:

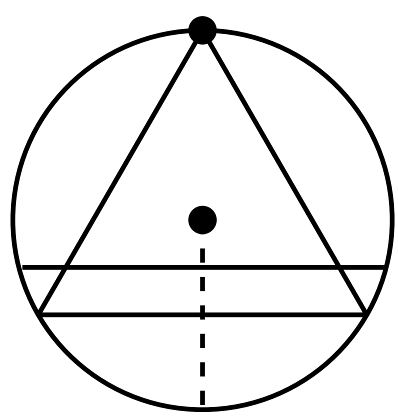
Problem. Take a circle of radius 1, and randomly choose a chord. What is the probability that this chord is $\geq \sqrt{3}$?

Proof. Here's one model of randomness: pick two points at random on the boundary of our unit circle, and draw the line connecting them. By rotating so that the first point is at the exact top of our circle, you can think of this as picking just one point somewhere on the perimeter of the circle below, and drawing a chord from this point to the point at the top of our circle.



You can check (geometry!) that the side lengths of the equilateral triangle inscribed above has side length $\sqrt{3}$; therefore, our chord has length $\geq \sqrt{3}$ if and only if we pick our point in the bottom $1/3$ of our circle. In other words, our chord will be greater than $\sqrt{3}$ $1/3$ of the time.

Alternately, here's another model of randomness: randomly create a chord by randomly picking a point in our circle and treating it as the midpoint of our chord. By rotation, we can assume that this midpoint occurs on the dotted line below:



Half of the choices of this point will correspond to chords longer than the base of our triangle (i.e. $\geq \sqrt{3}$) and the other half will correspond to shorter chords: i.e. our chord will be $\geq \sqrt{3}$ $1/2$ of the time.

[1] Usually they will use variables instead of saying "bargle." Usually.

The fact that $1/3 \neq 1/2$ shoud convince you of why carefully defining your probability space is a good idea.

Throughout this class, this usually won't be a problem: we're typically going to work in the probability space $G_{n,1/2}$, which is the set of all graphs on n vertices together with the probability function that sets

$$\mathbb{P}(A) = \frac{|A|}{|X|} = \frac{|A|}{2^{\binom{n}{2}}}.$$

In other words, we weight all of our graphs as being equally likely.

2 $o(n)$ Notation

This section is really straightforward.

Definition. A function $f(n)$ is said to be $o(n)$ if

$$\lim_{n\to\infty} \frac{f(n)}{n} = 0.$$

Roughly speaking, this means that $f(n)$ "grows much slower" than n as n heads off to infinity. We can generalize this from n to other functions: a function $f(n)$ is said to be $o(g(n))$ if

$$\lim_{n\to\infty} \frac{f(n)}{g(n)} = 0.$$

Again, roughly speaking, this means that $f(n)$ "grows much slower" than $g(n)$.

Example. The function $n \mapsto \sin(n)$ is $o(n)$, because

$$\lim_{n\to\infty} \frac{\sin(n)}{n} = 0.$$

Similarly, $\sin(n)$ is also $o(n^2)$, because $\lim_{n\to\infty} \frac{\sin(n)}{n^2} = 0$, and is also also $o(\ln(n))$, because $\lim_{n\to\infty} \frac{\sin(n)}{\ln(n)} = 0$..

3 Connecting Linear Algebra to Graph Theory

The following definitions really merit a course on their own; however, we only have a third of a lecture. Keep in mind that there is a lot more going on here than you may think, and that these ideas really merit closer study when you leave this class!

Definition. Take any $n \times n$ matrix A. We say that a vector $\mathbf{v} \in \mathbb{R}^n$ is an eigenvector of A, and $\lambda_1 \in \mathbb{R}$ is an eigenvalue of A, if $A\mathbf{v} = \lambda_1 \cdot \mathbf{v}$.

Definition. Given a graph G with vertex set $\{1, \ldots n\}$, we define its **adjacency matrix** A_G as the following $n \times n$ matrix:

$$A = \{a_{ij} : a_{ij} = 1 \text{ if } \{i, j\} \in E(G), \text{ and } 0 \text{ otherwise.}\}$$

One nice property of these adjacency matrices is that they are **symmetric** across their main diagonal: i.e. $(i, j) = (j, i)$, for any cell (i, j). We note this because of the following theorem:

Theorem 1 *(The spectral theorem:) Any symmetric matrix A has n eigenvalues (counting multiplicity, i.e. the eigenvalue 1 could be repeated multiple times) and n corresponding eigenvectors, all of which can be chosen to be orthogonal[2] to each other.*

In practice, we're going to use this to simply say that any graph's adjacency matrix will have n eigenvalues corresponding to different eigenvectors. We care about this because (surprisingly enough) the adjacency matric and eigenvalues of a graph are intimately related to other graph properties, like the number of 4-cycles in a graph or the connectivity of the graph! We present two quick properties here, to illustrate the power in these ideas and give you a taste of what we're going to do with linear algebra occasionally in this class:

Theorem 2 *Suppose G is a graph with vertex set $\{1, \ldots n\}$ with adjacency matrix A. Then the (i, j)-th entry of A^k denotes the number of distinct walks of length k from i to j.*

Proof. First, consider A^2, whose entries we're claiming are connected to the number of walks of length 2 in our graph.

Think about a walk of length two between two vertices i, j in the following sense: any such walk must first connect i to some vertex v with an edge, and then connect v to j. In other words, if we want to count the total number of these walks, we would want to calculate the sum

$$\sum_{v=1}^{n} \text{isEdge}(i, v) \cdot \text{isEdge}(v, j).$$

But wait! We've defined these isEdge functions earlier – specifically, we defined the adjacency matrix A_G of G in such a way that $a_{ij} = 1$ whenever there is an edge from i to j, and 0 otherwise. So, in this notation, we have that the number of walks from i to j is just

$$\sum_{v=1}^{n} a_{iv} \cdot a_{vk},$$

which we can recognize as the dot product

$$\begin{bmatrix} a_{i1} & a_{i2} & \ldots & a_{in} \end{bmatrix} \cdot \begin{bmatrix} a_{1j} \\ a_{2j} \\ \ldots \\ a_{nj} \end{bmatrix}.$$

But this is just the dot product of the i-th row and the j-th row of A_G! So, we've just proven the following:

[2]Two vectors $\mathbf{v}, \mathbf{w}$ are orthogonal if their dot product $\mathbf{v} \cdot \mathbf{w} = (v_1 w_1, \ldots v_n, w_n)$ is the 0-vector $(0, 0, \ldots 0)$.

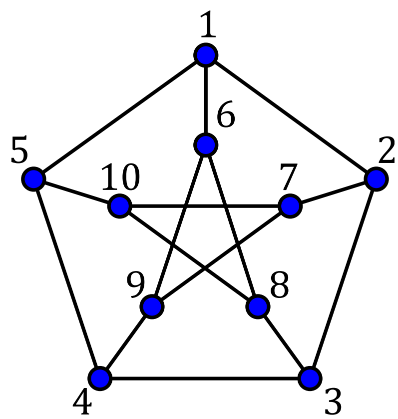
Proposition 3 *Suppose G is a graph with vertex set $\{1, \ldots n\}$ with adjacency matrix A. Then the (i, j)-th entry of A^2 denotes the number of walks of length 2 from i to j.*

We can easily generalize this to walks of length k, via induction on k. Suppose that we know that the entries of A^k correspond to the number of walks of length k from i to j. Given i and j, how can we find all of the walks of length $k+1$ from i and j?

Well: any walk of length $k+1$ from i to j can be described as a walk from i to some vertex v of length k, and then a walk of length 1 from v to j itself! So, if we just simply use the summation trick we used before, we can see that

$$\begin{aligned}
\text{numberOfWalks}_{k+1}(i,j) &= \sum_{v=1}^{n} \text{numberOfWalks}_k(i,v) \cdot \text{isEdge}(v,j) \\
&= (i,j) - \text{th entry of } A_G^k \cdot A_G \\
&= (i,j) - \text{th entry of } A_G^{k+1}.
\end{aligned}$$

The above hopefully illustrates why we would care about the adjacency matrix in this class: it gives us a fantastically useful way to calculate the number of paths in a matrix! In particular: take the Petersen graph



The adjacency matrix of this graph is

$$\begin{bmatrix}
0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\
0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0
\end{bmatrix}$$

You can calculate that the third power of this matrix is

$$\begin{bmatrix} 0 & 5 & 2 & 2 & 5 & 5 & 2 & 2 & 2 & 2 \\ 5 & 0 & 5 & 2 & 2 & 2 & 5 & 2 & 2 & 2 \\ 2 & 5 & 0 & 5 & 2 & 2 & 2 & 5 & 2 & 2 \\ 2 & 2 & 5 & 0 & 5 & 2 & 2 & 2 & 5 & 2 \\ 5 & 2 & 2 & 5 & 0 & 2 & 2 & 2 & 2 & 5 \\ 5 & 2 & 2 & 2 & 2 & 0 & 2 & 5 & 5 & 2 \\ 2 & 5 & 2 & 2 & 2 & 2 & 0 & 2 & 5 & 5 \\ 2 & 2 & 5 & 2 & 2 & 5 & 2 & 0 & 2 & 5 \\ 2 & 2 & 2 & 5 & 2 & 5 & 5 & 2 & 0 & 2 \\ 2 & 2 & 2 & 2 & 5 & 2 & 5 & 5 & 2 & 0 \end{bmatrix}$$

and the fourth power is

$$\begin{bmatrix} 15 & 4 & 9 & 9 & 4 & 4 & 9 & 9 & 9 & 9 \\ 4 & 15 & 4 & 9 & 9 & 9 & 4 & 9 & 9 & 9 \\ 9 & 4 & 15 & 4 & 9 & 9 & 9 & 4 & 9 & 9 \\ 9 & 9 & 4 & 15 & 4 & 9 & 9 & 9 & 4 & 9 \\ 4 & 9 & 9 & 4 & 15 & 9 & 9 & 9 & 9 & 4 \\ 4 & 9 & 9 & 9 & 9 & 15 & 9 & 4 & 4 & 9 \\ 9 & 4 & 9 & 9 & 9 & 9 & 15 & 9 & 4 & 4 \\ 9 & 9 & 4 & 9 & 9 & 4 & 9 & 15 & 9 & 4 \\ 9 & 9 & 9 & 4 & 9 & 4 & 4 & 9 & 15 & 9 \\ 9 & 9 & 9 & 9 & 4 & 9 & 4 & 4 & 9 & 15 \end{bmatrix}.$$

Using these observations, you can instantly see that (because the trace of A^3 is 0) there are no triangles as subgraph of the Petersen graph, because entries on the diagonal correspond to paths of length 3 that start and end at the same place: i.e. triangles! More obscurely, you can use that because every entry on the diagonal of A^4 is 15, that there are no 4-cycles in the Petersen graph (why? Prove this!)

Proposition 4 *A graph G is bipartite if its spectrum is symmetric about 0 (i.e. λ is an eigenvalue of A_G iff $-\lambda$ is.)*

Proof. Write $G = (V_1 \cup V_2, E)$,, where $V_1 = \{1, 2, \ldots k\}$ and $V_2 = \{k+1, k+2, \ldots n\}$ partition G's vertices. In this form, we know that the only edges in our graph are from V_1 to V_2. Consequently, this means that A_G is of the form

$$\begin{array}{|c|c|} \hline 0 & B \\ \hline B^T & 0 \\ \hline \end{array},$$

where the upper-left hand 0 is a $k \times k$ matrix, the lower-right hand 0 is a $n-k \times n-k$ matrix, B is a $(n-(k+1)) \times k$ matrix, and B^T is the transpose of this matrix.

Choose any eigenvalue λ and any eigenvector $(v_1, \ldots v_k, w_{k+1}, \ldots w_n) = (\mathbf{v}, \mathbf{w})$. Then, we have

$$A_G \cdot (\mathbf{v}, \mathbf{w}) = \begin{array}{|c|c|} \hline 0 & B \\ \hline B^T & 0 \\ \hline \end{array} \cdot \begin{bmatrix} \mathbf{v} \\ \mathbf{w} \end{bmatrix} = \begin{bmatrix} B \cdot \mathbf{w} \\ B^T \cdot \mathbf{v} \end{bmatrix} = \begin{bmatrix} \lambda \cdot \mathbf{v} \\ \lambda \cdot \mathbf{w} \end{bmatrix} = \lambda \begin{bmatrix} \mathbf{v} \\ \mathbf{w} \end{bmatrix}$$

But! This is not the only eigenvector we can make out of $\mathbf{v}$ and $\mathbf{w}$. Specifically, notice that if we multiply A_G by the vector $(\mathbf{v}, -\mathbf{w})$, we get

$$A_G \cdot (\mathbf{v}, -\mathbf{w}) = \begin{array}{|c|c|} \hline 0 & B \\ \hline B^T & 0 \\ \hline \end{array} \cdot \begin{bmatrix} \mathbf{v} \\ -\mathbf{w} \end{bmatrix} = \begin{bmatrix} -B \cdot \mathbf{w} \\ B^T \cdot \mathbf{v} \end{bmatrix} = \begin{bmatrix} -\lambda \cdot \mathbf{v} \\ \lambda \cdot \mathbf{w} \end{bmatrix} = -\lambda \begin{bmatrix} \mathbf{v} \\ -\mathbf{w} \end{bmatrix}.$$

In other words, whenever λ is an eigenvalue of A_G, $-\lambda$ is as well!